5

10

CLAIMS

What is claimed is:

1. Computer apparatus for gene prediction comprising:

a plurality of units for predicting gene locations in a subject genomic sequence, each unit providing respective intermediate indications of gene locations output by the unit;

a combiner coupled to each unit to receive the respective intermediate output indications of predicted gene locations, the combiner being formed of a Bayesian network which combines the predicted gene locations using probabilities of gene locations of the subject genomic sequence, and forming a final combined output for indicating predicted genes of the subject genomic sequence.

- 2. Computer apparatus as claimed in Claim 1 wherein the plurality of units is a plurality of expert systems.
- 15 3. Computer apparatus as claimed in Claim 1 wherein the Bayesian network accounts for dependencies between individual units and dependencies between adjacent parts of the subject genomic sequence.
 - 4. Computer apparatus as claimed in Claim 3 wherein the Bayesian network combines the predicted gene locations according to

20
$$Y^* = \max_{Y_t} P(Y_t | E_1, ..., E_n, Y^*_{t-1})$$

where t is location in the subject genomic sequence and $E_1,...,E_n$ are the respective predicted gene locations of individual units 1 through n, n being the number of units in the plurality.

15

- 5. Computer apparatus as claimed in Claim 1 wherein the subject genomic sequence is a DNA or RNA sequence.
- 6. Computer apparatus as claimed in Claim 1 wherein gene locations include exon predictions.
- 5 7. Computer apparatus as claimed in Claim 6 wherein gene locations further include exon and intron predictions; and

the final combined output indicates exons and introns of the predicted genes of the subject genomic sequence.

- 8. Computer apparatus as claimed in Claim 1 wherein the Bayesian network is trained on known genes of a known genome sequence.
 - 9. A method for automated gene prediction, comprising the steps of:

using a plurality of expert systems, preliminarily predicting gene locations in a subject genomic sequence;

using a Bayesian network, combining the preliminarily predicted gene locations to form a final combined output indicating predicted genes of the subject genomic sequence, the Bayesian network accounting for dependencies between individual expert systems and dependencies between adjacent parts of the subject genomic sequence.

A method as claimed in Claim 9 wherein the step of combining using a Bayesian
network combines according to

$$Y^* = \max_{Y_t} P(Y_t | E_1, ..., E_n, Y^*_{t-1})$$

where t is location in the subject genomic sequence and $E_1,...,E_n$ are the respective predicted gene locations of individual expert systems, n being the number of expert systems in the plurality.

- 11. A method as claimed in Claim 9 wherein the subject genomic sequence is a DNA or RNA sequence.
- 12. A method as claimed in Claim 9 wherein gene locations include exon predictions.
- 5 13. A method as claimed in Claim 12 wherein gene locations further include exon and intron predictions; and

the final combined output indicates exons and introns of the predicted genes of the subject genomic sequence.